Reducing Affective Responses to Surgical Images through Color Manipulation and Stylization

Lonni Besançon, Amir Semmo, David J. Biau, Bruno Frachet, Virginie Pineau, El Hadi Sariali, Rabah Taouachi, Tobias Isenberg, Pierre Dragicevic

Reducing Affective Responses to Surgical Images through Color Manipulation and Stylization

Lonni Besançon*
Inria & Université Paris-Saclay, France

Amir Semmo
Hasso Plattner Institute, University of Potsdam, Germany

David Biau
Assistance Publique – Hôpitaux de Paris, France

Bruno Frachet
Assistance Publique – Hôpitaux de Paris, France

Virginie Pineau
Institut Curie, France

El Hadi Sariali
Assistance Publique – Hôpitaux de Paris, France

Rabah Taouachi
Institut Curie, France

Tobias Isenberg
Inria & Université Paris-Saclay, France

Pierre Dragicevic
Inria & Université Paris-Saclay, France

Original photo of a lasagna dish. Output of HueShift2. Output of FlowAbs [Kyprianidis and Döllner 2008].

Figure 1: Two techniques studied in this article, each using a different strategy for making surgery images easier to look at.

ABSTRACT

We present the first empirical study on using color manipulation and stylization to make surgery images more palatable. While aversion to such images is natural, it limits many people's ability to satisfy their curiosity, educate themselves, and make informed decisions. We selected a diverse set of image processing techniques, and tested them both on surgeons and lay people. While many artistic methods were found unusable by surgeons, edge-preserving image smoothing gave good results both in terms of preserving information (as judged by surgeons) and reducing repulsiveness (as judged by lay people). Color manipulation turned out to be not as effective.

CCS CONCEPTS

• **Computing methodologies → Non-photorealistic rendering**; *Image processing*; • **Human-centered computing** → *Empirical studies in HCI*;

*Lonni Besançon is also with Linköping University, Sweden.

Expressive '18, August 17–19, 2018, Victoria, BC, Canada

ACM ISBN 978-1-4503-5892-7/18/08…$15.00
https://doi.org/10.1145/3229147.3229158

KEYWORDS

Stylization, affect, empirical study, surgery.

ACM Reference Format:
Lonni Besançon, Amir Semmo, David Biau, Bruno Frachet, Virginie Pineau, El Hadi Sariali, Rabah Taouachi, Tobias Isenberg, and Pierre Dragicevic. 2018. Reducing Affective Responses to Surgical Images through Color Manipulation and Stylization. In *Expressive '18: The Joint Symposium on Computational Aesthetics and Sketch Based Interfaces and Modeling and Non-Photorealistic Animation and Rendering, August 17–19, 2018, Victoria, BC, Canada.* ACM, New York, NY, USA, Article 4, 13 pages. https://doi.org/10.1145/3229147.3229158

1 INTRODUCTION

Many non-photorealistic and expressive rendering techniques deal with the stylization of 2D images or videos [Kyprianidis et al. 2013; Rosin and Collomosse 2013]. While much of this work was initially motivated by the desire to replicate artistic techniques and was only guided by a subjective visual comparison to existing artwork, researchers have begun to empirically evaluate the effects of stylization [Gooch et al. 2010; Isenberg 2013; Salesin 2002]. Some researchers argue, however, that controlled experiments are difficult in the context of expressive rendering [Mould 2014], and that we should rather concentrate on subjective evaluation [Mould 2014] and on the appreciation of resulting graphics [Hall and Lehmann

2013]. While such forms of evaluation arguably have their place in the context of the often art-inspired field of expressive rendering, the goal of creating expressive graphics is increasingly understood to incorporate more than the *"support of artists (or illustrators)"* or the *"creation of tools for visual expression for non-artists"* and to include, e. g., also *"illustrations [...] to inform [...] patients"* in a medical context [Isenberg 2016]. In this latter case it is then essential that we understand how different stylistic filters are perceived and experienced by real people, and that we thus study them empirically, through controlled experiments [Gooch et al. 2010].

In the past, in fact, some researchers have already examined such effects of stylization. Mandryk et al. [2011] resp. Mould et al. [2012], for instance, studied people's emotional response to stylized images and found that emotional responses was generally muted, and that responses concentrated around neutral feelings. Others [Dragicevic et al. 2013] argue (motivated by the results of earlier studies [Schumann et al. 1996]) that stylization may affect people's attitude toward a data visualization and result in longer times they spend looking at the visuals. Here, however, we are less interested in potentially positive effects of stylization, but instead in how much it can diminish negative affects caused by unpleasant pictures. Such pictures are involved when surgeons inform their patients before surgical procedures—because many people find surgery pictures repellent (e.g., [Sawchuk et al. 2002; Tolin et al. 1997]), effective communication can suffer. This context would seem like an ideal application case for expressive rendering [Isenberg 2016]. Nonetheless, the creation of effective illustrative visualizations of a wide variety of surgical procedures is still beyond our abilities. We thus study whether it is possible to use existing stylization techniques for 2D images—applied to real surgical images—to achieve a similar effect, and diminish the negative affect that surgery pictures can elicit. Applications go beyond patient information and include student training, media communication, and public education.

Below we discuss various image filtering and stylization techniques that can be used to dampen the negative affect elicited by surgery pictures. We then report on an interview session with four surgeons, who helped us differentiate between techniques that can preserve important information, and techniques that are unusable because they obfuscate too much. We then report on an experiment where the most promising techniques were tested on ordinary subjects. We found that all techniques can reduce the repulsiveness of surgery pictures as judged by participants, although spatial-domain techniques appear to be more potent than color manipulations. We conclude by a discussion and opportunities for future work.

2 BACKGROUND

In this section, we review related work in non-photorealistic rendering, before reviewing work from other areas on how people perceive, experience, and deal with surgery and injury images.

2.1 Non-photorealistic Surgery Illustrations

Medical illustration has long been one of the primary motivations for non-photorealistic and expressive rendering [Gooch and Gooch 2001; Strothotte and Schlechtweg 2002] and, consequently, many researchers have developed rendering techniques for this purpose. Several surveys and tutorials have covered the field in detail (e. g., [Costa Sousa et al. 2005; Ebert and Costa Sousa 2006; Preim and Botha 2014; Viola et al. 2006]) and we refrain here from citing specific techniques. Common among them is, however, that they are inspired by traditional, usually hand-made illustration techniques, styles, and examples and that they thus focus on clarification and explanation, rather than on emotional aspects or on reducing the negative affect that some types of content could induce in people.

Another common characteristic of many illustrative techniques and also traditional illustration styles—for medical application and otherwise—is the use of abstraction and emphasis. These aspects have been discussed in the visualization and expressive rendering literature, such as in the contributions by Rautek et al. [2008] and Viola and Isenberg [2018]. Here, abstraction is *"a transformation which preserves one or more key concepts and removes detail that can be attributed to natural variation, noise, or other aspects that one intentionally wants to disregard from consideration"* [Viola and Isenberg 2018]—to allow viewers of a visualization to focus on major or important aspects. In this work, however, we explore the abstracting qualities of image filters for the removal of details such that the images are perceived as less offensive—potentially because they no longer depict surgery situations in all their details.

2.2 Non-photorealistic Techniques and Affect

In the past, researchers have studied how stylization can influence how people perceive images. Gooch and Willemsen [2002], for example, showed that a line-based rendering of a virtual scene leads participants to underestimate distances by about a third, quite similar to what happens in 'photorealistic' VR settings. Later, Gooch et al. [2004] showed that non-photorealistic illustrations and caricatures of people's portraits could be learned faster than real photographs. We cannot deduct from these results, however, that stylized images would lead people to feel differently about what is shown.

Already early work on non-photorealistic rendering, however, discussed this very effect. Duke et al. [2003] and Halper et al. [2003], for example, described how the (non-photorealistic) depiction style can affect people's assessment of danger and safety as well as strength and weakness, and can change their participation and interaction behavior (for study details see Section 2 of Halper's [2003] thesis). Even before this work, Schumann et al. [1996] provided evidence for stylization to increase people's willingness to interact with visuals. More recently, McDonnell et al. [2012] showed that an increased abstraction of virtual characters (according to their participants' classification of "realism") decreases appeal, friendliness, and trustworthiness up to a point; for highly abstracted depictions people again feel similar about the stylized virtual characters as they do for realistic depictions—similar to what the Uncanny Valley theory predicts. Like the perceptual studies discussed before, however, these approaches do not shed light on whether stylization changes people's negative emotions toward disturbing images.

Most relevant for our own work, out of the expressive rendering literature, is Mandryk et al.'s [2011] and Mould et al.'s [2012] work who demonstrated that stylization can affect the emotional interpretation of images. Similar to what we do in our experiment, they applied a range of styles (stippling, line art, painterly rendering, and blur) to a set of images with different affective content from the International Affective Picture System (IAPS), and analyzed people's feeling of arousal, valence, dominance, and aesthetics. Stylization generally muted participants' emotional responses toward a neutral point, yet emotions were never completely suppressed. Their negative stimuli (e. g., a gun pointed at camera, or a cemetery), however, did not have the repulsive potency that surgery photos can have. This study thus inspires our own, but we specifically target surgery pictures that many people cannot easily look at.

We note that researchers also have examined the opposite path: changing the stylization of images based on emotions detected in a video feed. Shugrina et al. [2006], for example, presented their "empathic painting" technique that recognizes a person's emotional state based on features of their facial expression, which they then use to adjust the parameters of a painterly rendering technique. Here Shugrina et al. borrow from the psychological literature and created a mapping from the detected emotional state to rendering parameters such as stroke path and color. Yet, it is not clear if the resulting images also change the emotional state of the viewer, or if so in what way this process can be controlled.

2.3 Human Response to Surgery Imagery

For several decades, researchers have studied human response to repellent images in order to uncover the physiological and psychological mechanisms involved. Studies have used various types of aversive stimuli such as homicide scenes [Hare et al. 1970], spiders [Tolin et al. 1997], vomit [Olatunji et al. 2008], maggots, cadavers, and dirty toilets [Schienle et al. 2002]. Many studies have examined responses to scenes depicting a body envelope violated by an injury or a surgery. Examples include photos of body mutilation (e.g., [Klorman et al. 1977]), of surgery procedures (e.g., [Sawchuk et al. 2002; Tolin et al. 1997]), and videos of medical interventions such as blood draw [Gilchrist and Ditto 2012], open-heart surgery [Olatunji et al. 2008], or surgical amputation [Rohrmann and Hopp 2008]. Studies have involved both ordinary subjects (e.g., [Hare et al. 1970]), BII-phobic[1] subjects (e.g., [Öst et al. 1984]), and often a combination of both (e.g., [Haberkamp and Schmidt 2014]).

Various measurements have been employed to quantify subject reactions, the most common being heart rate (e.g., [Klorman et al. 1977; Olatunji et al. 2008]). Others include facial expression (using videotaping [Lumley and Melamed 1992] or electromyography [Lang et al. 1993; Olatunji et al. 2008]), skin conductance [Lang et al. 1993], neural activation using fMRI [Schienle et al. 2002], attentional avoidance using eye tracking [Armstrong et al. 2013], and visuomotor processing using a response priming task [Haberkamp and Schmidt 2014]. Subjective measures were also used, where subjects were asked to report to what extent they felt fear and disgust [Sawchuk et al. 2002; Tolin et al. 1997], avoided watching [Olatunji et al. 2008], or experienced vasovagal (i.e., pre-fainting) symptoms [Gilchrist and Ditto 2012]. A strong reaction to a body injury depiction is often marked by a decrease in heart rate, or an increase followed by a rapid decrease called "diphasic response" [Cisler et al. 2009]. It also often involves activation of the corrugator supercilii (the "frowning muscle") and the levator labii (responsible for lifting the upper lip) [Cisler et al. 2009]. However, studies are inconsistent and there appears to be no perfectly reliable measure that can consistently elicit the same response [Cisler et al. 2009].

Most of these studies were conducted in order to untangle the emotions involved when people witness surgeries or injuries, sometimes in the hope of better treating BII phobia. This has proven hard to study, as reactions seem to involve various emotions such as anxiety, fear, disgust, and vicarious pain [Benuzzi et al. 2008; Cisler et al. 2009]. In particular, the relative role of fear vs. disgust has long been a subject of debate, although now the consensus seems to be that disgust is the main emotion involved [Cisler et al. 2009; Olatunji et al. 2010]. To understand why, it helps to recall that fear has evolved for organisms to run away from threats such as spiders, but for static content like body injuries, no such response is necessary [Cisler et al. 2009]. More likely, body injuries are experienced as repellent in order to prevent risks of disease or contagion following physical contact, which requires a disgust response [Cisler et al. 2009]. Chapman and Anderson [2012] introduced a taxonomy of disgust where *blood-injury disgust* is a subtype of *physical disgust*, and whose evolutionary function is to avoid infection. Olatunji [2008], however, distinguishes *contamination disgust* from *animal-reminder disgust*, with animal-reminder disgust being elicited by *"attitudes and practices surrounding sex, injury to the body or violations of its outer envelope, and death"* which all act as *"reminders of our own mortality and inherent animalistic nature"* [Olatunji et al. 2008].

Despite all this previous work, human reactions to the sight of surgery scenes remains poorly understood. Our goal is not to further this understanding, but simply to find out whether processing surgery photos can dampen their affective potency. As far as we know, all studies on blood-injury disgust have either assessed aversive stimuli in isolation or compared them with neutral stimuli, and none of them has studied the effect of processing aversive stimuli using filters or stylization. When conducting our study, we drew from the experience accumulated in this research area, but simplified the methods to directly answer our research question.

In parallel to this body of work focusing on blood-injury disgust, there has been work in psychology and neurosciences where various types of emotionally-salient stimuli were used to study emotion and cognition. Such stimuli were used, for example, to study cultural differences in emotion processing [Wrase et al. 2003], and emotion regulation [Eippert et al. 2007]. Some of the stimuli involved surgery and injury photos but again, affective neutralization through image processing has not been a focus. Nevertheless, this area of research has produced standardized stimuli sets which we will use for our own study, as explained in Section 5.1.

2.4 Picture Censorship Practices

On a societal level, offensive imagery has been addressed in two major ways: legal censorship and *de facto* (or self) censorship. While there appears to be close to no legal restriction on what visual content can be published in newspapers [Tooth 2014] or in Wikipedia [Wikipedia Contributors 2010a,b], films and videos games are usually regulated by rating systems to classify the media with regard to its suitability for different audiences. While movies cannot be easily customized, the video game industry has explored a wide range of "adjustable censorship" techniques. Some old video games had violent and sexual content disabled by default, while giving the option to reactivate it through the use of secret codes. More

[1] Blood Injection and Injury phobia (BII phobia) refers to *"an extreme and irrational fear of blood, injuries, or of receiving an injection or an invasive medical procedure"* which affects about 3.5% of the population [Haberkamp and Schmidt 2014].

elaborate adjustable censorship techniques were also developed: some video games (e.g., Silent Hill, Resident Evil, House Of The Dead, later release of Ocarina of Time) offer the option to change the color of blood to various tones such as blue, dark, or green depending on the game [TVTropes Contributors 2018a; ZeldaWiki Contributors 2018]. While most mangas feature black blood due to the constraints of black & white printing, some color animes and animated films employ a different blood color to suit all audiences (e.g., Dragon Ball Kai, Pokémon, Bleach, The little mermaid) [TVTropes Contributors 2018b]. Similarly, in movies, black and white has been occasionally used to censor scenes with excessive bloodshed [Kill Bill Wiki Contributors 2017]. All such practices suggest that blood is considered to epitomize violence, but once deprived from its characteristic red color it seems to suddenly become inoffensive in people’s minds. These practices provide motivation for considering simple color manipulation techniques in our work.

3 CHOICE OF PROCESSING TECHNIQUES

In this article we use the term *image processing technique* or simply *technique* to refer to any procedure that transforms an image into another image, while keeping it recognizable. We considered four classes of techniques of varying complexity: color manipulation, edge-preserving smoothing, edge detection and enhancement, and image-based artistic rendering. We first outline relevant work and provide rationales for the techniques we retained, and then provide parameter settings yielding reasonable levels of abstraction. The 13 techniques and their settings are summarized in Table 1 and illustrated in Figure 1 and Figure 2 with a 1024x768 photo.[2]

3.1 Color Manipulation

We considered two approaches for changing blood color: decolorization and recolorization [Pratt 2007].

Decolorization. Grayscale conversion is a popular method for image decolorization, where the main challenge is to preserve and make use of the chrominance components so that perceptual image features are retained [Čadík 2008; Ma et al. 2015]. Most algorithms transform the problem into optimization to preserve salient features, e. g., by quantifying color differences between image locations [Gooch et al. 2005] or prevailing chromatic contrasts [Grundland and Dodgson 2007], optimizing color and luminance contrasts [Neumann et al. 2007], or considering the Helmholtz-Kohlrausch color appearance effect [Smith et al. 2008]. The latter localized apparent greyscale algorithm performed best in a previous experiment [Čadík 2008] and we thus retained it and named it ApparentGrey in this article. However, the method may suffer non-homogeneity artifacts near region boundaries, which can be addressed with a global mapping scheme [Kim et al. 2009].

Recolorization. A simple yet effective method of recolorization is to alter all color chrominances by a hue shift in the HSV space. We consider a uniform hue shift which makes blood green to which we refer to as HueShift. The hue shift shown in Figure 1 uses different settings and we discuss it in Section 5.2. A more sophisticated approach could involve color transfer between source and target images or color palettes, which typically relies on image statistics to globally and locally control color distributions [Faridul et al. 2014].

[2]Our custom image processing software is available at https://osf.io/4pfes/ .

3.2 Edge-preserving Image Smoothing

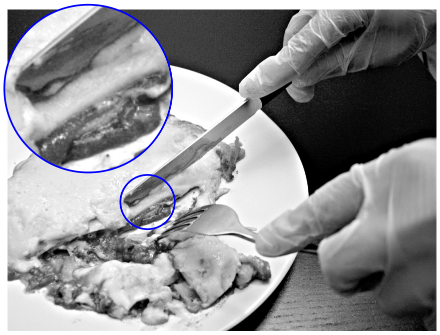
While color manipulation might reduce the emotional impact carried by blood, it preserves the details of the original photo. A black-and-white photo, in particular, may still appear too crude. Therefore, we consider other types of filters, starting with edge-aware image smoothing as a building block for abstraction, artistic stylization and tone mapping. Numerous filter-based techniques have been proposed to approach these applications in an automated way, most of them deriving local image structures for feature-aware processing.

Bilateral filter. The bilteral filter is a popular choice to approximate an anisotropic diffusion, which works by weight averaging pixel colors in a local neighborhood based on their distances in space and range [Tomasi and Manduchi 1998]. It weights pixels with a high difference in intensity less than a Gaussian filter to preserve image structures at a better scale. We retain it and refer to it as Bilateral. Most relevant applications apply the bilateral filter in a multi-stage process for real-time rendering with a cartoon look [Winnemöller et al. 2006], and enhance it by flow-based implementations adapted to the local image structure [Kang et al. 2009; Kyprianidis and Döllner 2008]. Providing smooth outputs at curved boundaries of delicate structures, we thus consider the flow-based variant [Kyprianidis and Döllner 2008], and name it FlowAbs. As a generalized variant, the guided filter [He et al. 2013] may provide similar characteristics with reduced unwanted gradient reversal artifacts, but only provides a non-feature-aligned implementation.

Mean-shift. A mean-shift is a popular approach for edge-preserving smoothing [Comaniciu et al. 2002] and saliency-guided image abstraction [DeCarlo and Santella 2002]. It provides a non-parametric filter that estimates probability density functions by iteratively shifting color values to averaged color values of a local neighborhood. However, the approach typically produces rough boundaries that is more suited to image segmentation.

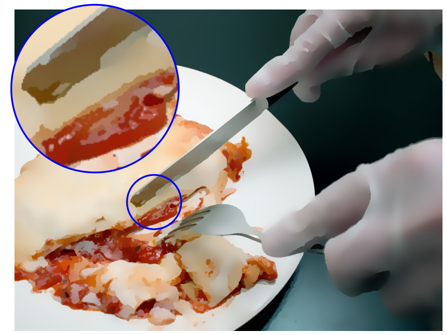
Kuwahara filter. A popular approach that works accurately even with high-contrast images—contrary to the bilateral filter—and provides smoothed outputs at curved boundaries, is the Kuwahara filter [Kuwahara et al. 1976] and its generalized [Papari et al. 2007] and anisotropic [Kyprianidis 2011; Kyprianidis et al. 2009] variants. The kernel of the anisotropic Kuwahara filter is divided into shape-aligned overlapping subregions, where the response is defined as the mean of the subregion with minimal variance. We retain the multi-scale variant [Kyprianidis 2011] and refer to it as Kuwahara. It maintains a uniform level of abstraction due to local area flattening and can scale with the image resolution.

Shock, morphological, and geodesic filters. Contrary to previous filters, additional categories weight colors across feature boundaries for higher levels of abstraction, for which we retain methods with shock filtering, i. e., in conjunction with a constrained mean curvature flow [Kang and Lee 2008] (ShapeSimpl) and diffusion tensors for coherence-enhancing abstraction [Kyprianidis and Kang 2011] (CoherenceEnh). Morphological filtering based on dilation and erosion, and geodesic filtering using distance transforms are also popular choices to obtain results of high abstraction [Criminisi et al. 2010; Mould 2012], but were found to require local control to effectively adjust the level of abstraction.

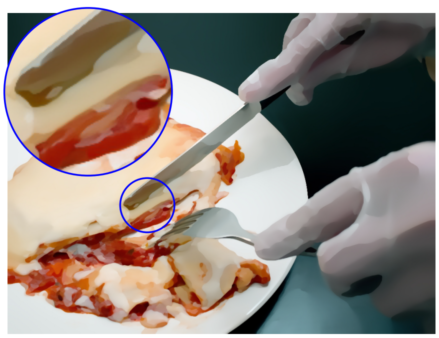


ApparentGrey [Smith et al. 2008]

HueShift

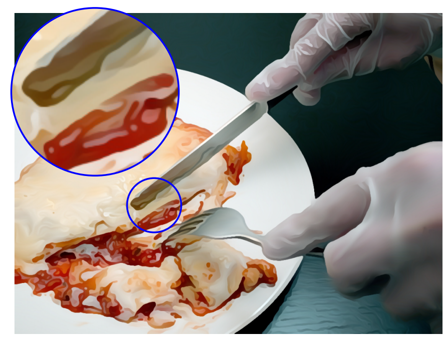


Bilateral [Tomasi and Manduchi 1998]

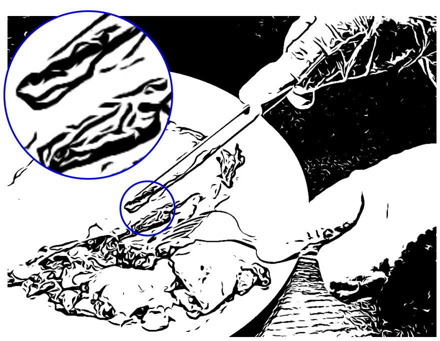


Kuwahara [Kyprianidis 2011]

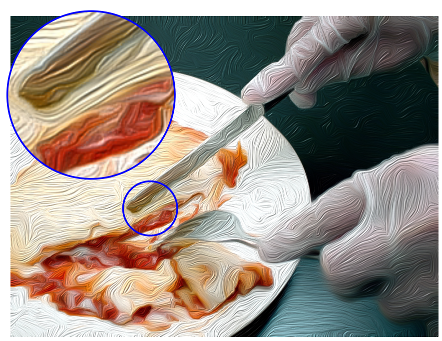
ShapeSimpl [Kang and Lee 2008]



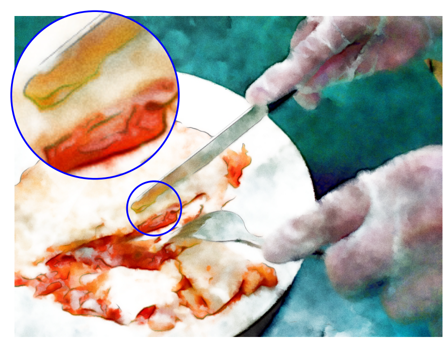
CoherenceEnh [Kyprianidis and Kang 2011]



XDoG [Winnemöller et al. 2012]

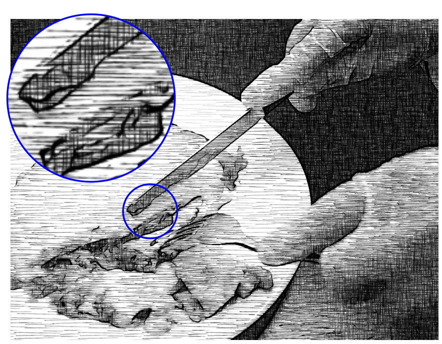


Oilpaint [Semmo et al. 2016]

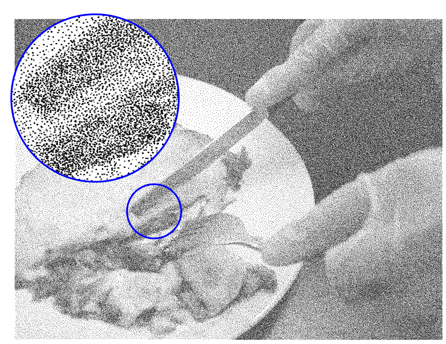


Watercolor [Bousseau et al. 2006; Wang et al. 2014]

BrushStrokes [Hertzmann 1998]



Hatching [Praun et al. 2001]



Stippling [Martín et al. 2010; Martín et al. 2011]

Figure 2: Image processing techniques used in the interviews with surgeons, together with FlowAbs shown in Figure 1.

Table 1: Overview of techniques with their correspondent parameters.

	Abbreviation	Based on works	Parameter settings
Filters	ApparentGrey	[Smith et al. 2008]	$N = 4$, $p_i = 0.5$, $k_i = 0.5$
	HueShift	custom, see Section 3.5	$\alpha = 120.0$
	Bilateral	[Tomasi and Manduchi 1998]	$\sigma_d = 3.0$, $\sigma_r = 4.25\%$
	FlowAbs	[Kyprianidis and Döllner 2008]	$\rho = 2$, $n_e = 1$, $n_a = 3$, $\sigma_d = 6$, $\sigma_r = 5.25\%$, $\sigma_e = 1$, $\tau = 0.99$, $\epsilon = 0$, $\varphi_e = 2$, $\sigma_m = 3$, $q = 8$, $\varphi_q = 2$
	Kuwahara	[Kyprianidis 2011]	$N = 8$, $\rho = 2.0$, r = 6, $q = 8$, $\alpha = 1$, $\tau_w = 0.02$, $p_s = 0.5$, $p_d = 1.25$, $\tau_v = 0.1$
	ShapeSimpl	[Kang and Lee 2008]	$N = 8$, $k = 8$, $r = 1$, $\text{etf}_{\text{halfw}} = 3$, $\text{etf}_N = 4$, $\sigma_{\text{shock}} = 1$, $\tau_{\text{shock}} = 1$
	CoherenceEnh	[Kyprianidis and Kang 2011]	$N = 4$, $\sigma_d = 1$, $\tau_r = 0.002$, $\sigma_t = 6$, $\sigma_i = 0$, $\sigma_g = 1.5$, $r = 2$, $\tau_s = 0.005$, $\sigma_a = 1.5$
	XDoG	[Winnemöller et al. 2012]	$\sigma_c = 2.28$, $\sigma_e = 1$, $\sigma_m = 4.4$, $p = 99.0$, $\varphi = 100.0$, $\epsilon = 0.65$, $\sigma_a = 1.0$
Stylization	Oilpaint	[Semmo et al. 2016]	$\sigma_s = 12.0$, $n_e = 0$, $\sigma_b = 12.0$, $k_{scale} = 5.0$, $k_{specular} = 0.8$, $k_{shininess} = 12.0$
	Watercolor	[Bousseau et al. 2006; Wang et al. 2014]	Effects of [Bousseau et al. 2006; Wang et al. 2014] using FlowAbs: $\rho = 2.0$, $\sigma_d = 4.0$, $\sigma_r = 15.00\%$
	BrushStrokes	[Hertzmann 1998]	$T = 200$, $\text{size}_{\max} = 8$, $f_c = 1$, $f_\sigma = 0.5$, $j_{\{h,s,v\}} = 0$, $j_{\{r,g,b\}} = 0.3$, $f_g = 1$, $\hat{\sim} = 0.5$, length = [4, 16]
	Hatching	[Praun et al. 2001]	Art map of [Praun et al. 2001] scaled at 0.6, linear mapping with luminance, edge settings of FlowAbs
	Stippling	[Martín et al. 2011]	$res_0 = 1200$ ppi, $f_p = 2.0$, placement randomness = 25%, distribution = *normal*, colors = *b&w*, $\tau = 127$

Filters using global optimizations. Many filters focus on image decompositions by solving optimization problems to separate detail from base information, e. g., based on weighted least squares [Farbman et al. 2008], histograms [Kass and Solomon 2010], and gradient minimization [Xu et al. 2011]. However, they are found to have strengths in applications requiring complementary global optimizations, such as tone mapping and colorization. Moreover, they are typically not suited for interactive applications.

3.3 Edge Detection and Enhancement

Winnemöller et al. [2012] distinguish between gradient-based edge detection that thresholds the gradient magnitude of an image and Laplacian-based edge detection that identifies zero-crossings in the second derivative. Popular gradient-domain approaches identify image gradients with high magnitudes by using convolution filters, such as the Prewitt and Sobel filter [Pratt 2007], with subsequent thresholding of the magnitude. The approach is popular with medical images to ease object recognition, however produces results that are sensitive to noise. The Canny edge detector [Canny 1986] as a multi-stage algorithm provides several enhancements by combining smoothing and differentiation operators. However, although popular with MRI and CT images, it is more directed to semantic segmentation and may produce disconnected edge segments.

A real-time approach that is less sensitive to noise is to approximate the Laplacian of Gaussian [Marr and Hildreth 1980] using difference of Gaussians (DoG). The approach has shown to provide smooth edges of delicate structures, e. g., with respect to human faces [Gooch et al. 2004]. Therefore, we use the enhanced separable flow-based implementations of the DoG [Kang et al. 2007; Kyprianidis and Döllner 2008; Winnemöller et al. 2012] for FlowAbs, since they are adapted to the local orientation of an input image to create smooth coherent outputs for line and curve segments. We also retain the XDoG filter [Winnemöller et al. 2012] as a generalized approach that is able to obtain two tone black-and-white images, which relates to drawings found in illustrative visualization.

3.4 Image-based Artistic Rendering

Artistic image stylization has been suggested to dampen emotional responses [Mandryk et al. 2011]. We consider stylization techniques that simulate traditional media and painting techniques found in illustrative visualization, i.e., watercolor, oil paint, pen-and-ink, and stippling, for which we strive for classical or state-of-the-art methods that cover the taxonomy proposed by Kyprianidis et al. [2013]. Image filters are prominently used as building blocks of complex stylization effects, such as the bilateral filter and DoG to obtain toon renderings (FlowAbs), and flow-based Gaussian smoothing for more abstract filtering that simulates oil paint [Semmo et al. 2016] (Oilpaint). In addition, we use a Watercolor technique that simulates effects such as pigment density variation, edge darkening, wet-in-wet, and wobbling [Bousseau et al. 2006; Wang et al. 2014]. For stroke-based rendering, a popular method is to iteratively align brush strokes of varying color, size, and orientation according to the input image, for which we use the classical approach of Hertzmann [1998] (BrushStrokes). Techniques for tonal depiction typically direct tonal art maps based on luminance, for which we use a 2D hatching implementation that borrows from Praun et al. [2001] coupled with a DoG-based edge (Hatching). Finally, we consider the example-based stippling technique described by Martín et al. [2011] (Stippling), as it is able to provide scale-dependent results.

3.5 Parameters and Configurations

The parameters and configurations summarized in Table 1 are based on reported results and presets of the respective works, and experiments with medical images at a resolution of 1024×768 pixels.

Filters. We set the ApparentGrey filter to use default settings with uniform spatial control of four subbands to locally adjust local chromatic contrasts. For the HueShift, we use a shift of −120.0 degrees on the *hue* channel mapped to the HSL color wheel. We align the Bilateral filter to obtain a soft Gaussian smoothing with a spatial distance $\sigma_d = 3.0$ using additional filtering in the CIE-Lab color space using an increased weight ($\sigma_r = 4.25\%$). For FlowAbs, we use default parameters for edge enhancement with a doubled distance for bilateral filtering to compromise with the 512×512 pixel images used by Kyprianidis and Döllner [2008]. We set the Kuwahara filter to use a typical configuration with a radius of six pixels aligned to eight sectors, a slightly smoothed structure tensor, and multi-scale estimation [Kyprianidis 2011]. For deliberate smoothing across shape boundaries using ShapeSimpl and CoherenceEnh, we configure these filters to perform a single step of shock filtering after every iteration of mean curvature flow—i. e., four [Kyprianidis and Kang 2011]) and eight [Kang and Lee 2008] steps in total respectively. Finally, we set the edge enhancement using XDoG to output fine coherent lines with high details and a two-tone thresholding to sparsely obtain negative edges [Winnemöller et al. 2012].

Artistic Styles. Here we mainly seek to replicate the level of abstraction targeted by the respective works: For Oilpaint, smoothing parameters with a light paint texture that fall into the medium range as described by Semmo et al. [2016]; An implementation for Watercolor using flow-based bilateral filtering of FlowAbs but with wider filter kernels to achieve a similar level of abstraction as done by Bousseau et al. [2006] and Wang et al. [2014]; The "colorist wash" preset of BrushStrokes defined by Hertzmann [1998] to produce semi-transparent layered brush strokes; The default art map used by Praun et al. [2001] that is linearly mapped to the luminance in CIE-Lab color space; And default parameters for Stippling described by Martín et al. [2011] at highest resolution, with a normal distribution and black-and-white thresholding.

4 INTERVIEWS WITH SURGEONS

To find out which of these 13 techniques are usable in practice, we interviewed four surgeons: two otolaryngologists (respectively 13 and 35 years of experience), one orthopaedic surgeon, and one reconstructive surgeon (both 10 years of experience). Although all four surgeons are co-authors of this article, none of them was involved in the research at the time of the interviews.

We asked each surgeon to send us three of their own surgery photos that could help them explain a specific procedure to non-experts. We cropped (when necessary) and resized them to 1600×1200 pixels before processing them with the configurations detailed in Section 3.5. We then printed each processed image on a separate A4 sheet. At the beginning of each session, we asked the surgeon to compare and classify the processed images by making piles based on how useful the image would be as a support for communication and explanation, especially in terms of how much important information is preserved. Two images perceived as equally useful would go into the same pile. This procedure was repeated three times, once for each photo. Each surgeon saw 3 photos × 13 techniques = 39 images in addition to the 3 original photos. To limit ordering effects, we randomized the order in which processed images were presented. After the classification, we let surgeons further comment on the techniques and inquired them about possible applications.

4.1 Reported Preferences

We processed self-reported preferences as follows: first, for each combination of surgeon × photo, we assigned a number to each technique according to the pile it was in: 1 for the leftmost (most preferred) pile, 2 for the second pile, and so on. We then normalized these ranks using the halfway accumulative distribution [Jin and Si 2004]. This method gives each rank a score between 0 and 1 that corrects for possible differences in the way ranks are assigned (e.g., when some surgeons make more piles than others). We then derived *preference scores* by inverting the normalized ranks ($y = 1 - x$). All preference scores are shown in Figure 3. Finally, we averaged preference scores across pictures and surgeons to derive a single *aggregated preference score* per filter. These scores are also shown in Figure 3, on top. As we can see, CoherenceEnh was the most preferred technique, followed by FlowAbs, Kuwahara and Bilateral. We discuss the surgeons' comments in the following.

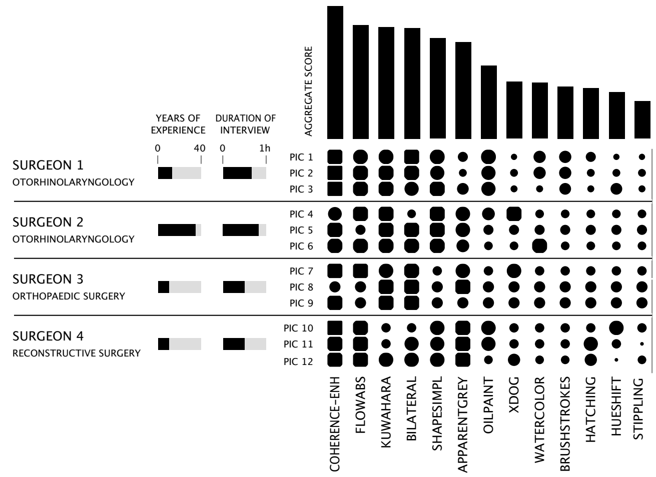


Figure 3: Tabular visualization [Perin et al. 2014] summarizing preference scores for each combination of surgeon, photo, and technique. Techniques are sorted according to their aggregate score, from most preferred to least preferred.

4.2 Qualitative feedback

We encouraged surgeons to voice comments both during the classification and in the debriefing interview.[3] In general, they reported that some of our processed images could be used for textbooks or classes (× 2 surgeons), to communicate with patients (× 2), and surprisingly even to communicate with other experts (× 1), as "drawings can be augmented" with notes for instance. One surgeon reported they could be particularly useful for plastic surgery and that it would be interesting to see how automatic processing could help communicate with children patients, as they are more sensitive to surgery images. We now report on the comments that were made for more than one photo and/or by more than one surgeon.

To begin with, the Bilateral technique was found to be usable for patients or in a book (× 2 photos, 1 surgeon). In addition to its good ranking, the FlowAbs technique's ressemblance to cartoons or comic strips was pointed out, and with it the possibility to remove a bit of "violence" from the photo (× 2, 1). It was also praised for the high visibility it gave to contours (× 3, 2). In contrast, the Hatching technique was reported to remove too many details (× 5, 3). Similarly, while HueShift only manipulated colors, it was reported to cause loss of information (× 3, 1), and made it hard to find anatomical correspondences (× 2, 1). The Oilpaint technique generated mixed reactions. On the one hand, it was praised for it artistic look (× 3, 2) and could potentially be used in books or with patients (× 2, 1). On the other hand, it was reported to remove useful information (× 2, 2). BrushStrokes was also reported to be artistic and possibly useful with patients (× 2, 2). The same qualities were reported for the ShapeSimpl technique. Stippling was explicitly reported as not usable (× 3, 1) and causing too much loss of information (× 7, 3). The Watercolor technique also removed too many details (× 5, 2). Finally, XDoG was also found to cause too much information loss (× 2, 1) as it makes it difficult to distinguish colors and contours (× 5, 2).

[3]Interview notes are available at osf.io/4pfes/.

Table 2: Techniques tested in the experiment.

Abbreviation	Parameter settings
ApparentGrey	Same as Table 1
HueShift2	$\alpha = -120.0$
ShapeSimpl2	Same as Table 1 except $N = 4$, $k = 4$
CoherenceEnh	Same as Table 1
Kuwahara	Same as Table 1
FlowAbs	Same as Table 1

5 EXPERIMENT WITH LAY PEOPLE

Our interview with surgeons helped us understand which processing techniques preserve key information from surgery photos. However, it is hard for surgeons to accurately predict the affective impact of processed photos on lay people. Thus, we conducted an experiment where we presented surgery photos to 30 participants, both unprocessed and processed, and asked them to rate them according to how repulsive they are. This experiment was approved by Inria's ethics committee (COERLE, approval number 2017-015).

5.1 Pictures

Throughout this section, we use the term *picture* to refer to an original photo, and the term *stimulus* to refer to a processed or unprocessed picture that is meant to be presented to participants.

We selected five pictures from two research catalogs: the Nencki Affective Picture System (NAPS) [Marchewka et al. 2014] and the International Affective Picture System (IAPS) [Lang et al. 1997]. These catalogs contain a range of emotionally-evocative photos that have been validated to elicit a positive, neutral, or negative affect. We selected five surgery photos among the negative pictures:[4]

- People 202 (NAPS)—a leg surgery,
- People 216 (NAPS)—a leg surgery or autopsy,
- People 221 (NAPS)—a surgery in the eye area,
- 3212 (IAPS)—a surgery performed on a dog, and
- 3213 (IAPS)—a finger surgery.

In addition to those five *surgery pictures*, we selected five *neutral pictures* from the NAPS catalog, all with consistent ratings of 1 on the disgust scale. These photos include, e.g., a surfer riding a wave and a man walking on the beach with his son.

Picture resolution is 1600×1200 for NAPS and 1024×768 for IAPS. For consistency, we rescaled all NAPS pictures to 1024×768. Thus, the effect of spatial filters will be slightly stronger than for the interview sessions, which used 1600×1200 pictures.

5.2 Techniques

We first selected the five most usable techniques according to surgeons (see Figure 3). We observed that Kuwahara and Bilateral yielded almost identical results on our experimental stimuli, and were therefore likely to elicit the same response. Thus, in order to save experimental conditions, we decided to remove Bilateral, since it is already used as a building block of FlowAbs. We further decided to add HueShift in order to cover a broader spectrum of approaches, even though it was ranked poorly. Looking back at surgeon pictures, we realized that HueShift often gives skin and blood the same green tones. This might be the reason why surgeons

[4]Usage restrictions do not allow us to reproduce or distribute the pictures. These can be obtained from the Nencki Institute of Experimental Biology and CSEA Media.

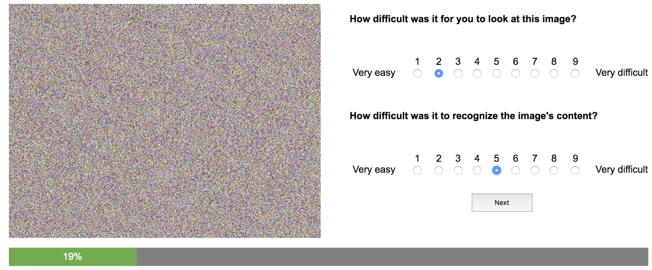


Figure 4: Experiment screen after stimulus presentation.

found that HueShift suppressed important information. Thus, we changed the shift angle from 120° to −120° in the HSL space, which maps blood and skin to blue/purple tones where color discrimination is superior [Bachy et al. 2012]. We refer to this technique as HueShift2. Finally, by examining results on experimental stimuli, we realized that the higher number of iterations used for ShapeSimpl eliminated significantly more details than other techniques. We therefore tuned the settings in order to get more comparable levels of abstraction. We refer to this technique as ShapeSimpl2. This leaves us with 6 techniques, summarized in Table 2.

5.3 Metrics

A variety of psychophysiological measures exist to quantify emotional response (see Section 2.3), but they tend to be noisy and they require specialized equipment. We thus decided to simply measure self-reported subjective experience. Psychology has adopted standardized scales to assess affect, such as valence, arousal and disgust scales [Lang et al. 1997; Marchewka et al. 2014]. However, the difficulty of looking at a repellent surgery image may not directly map to either valence, arousal or disgust as they are understood by participants. Therefore, we chose to ask a more direct question, i.e., *"how difficult was it for you to look at this image?"*, on a 9-point scale from *very easy* to *very difficult* (see Figure 4). The question was framed in the past tense because, as we will see later on, participants did not see the stimulus when they answered the question.

As a complement to the expert interview, we also asked participants to estimate to what extent the content of the scene has been obfuscated by the filter. The question was *"how difficult was it to recognize the image's content?"*, again on a 9-point scale from *very easy* to *very difficult.* The meaning of both questions was explained in preliminary instructions, with examples.

5.4 Design and Procedure

Because between-subject designs typically suffer from low statistical power [Bellemare et al. 2014], we used a within-subject design. Each participant saw all combinations of *picture* and *technique.* With 10 pictures (5 surgery, 5 neutral) and 7 techniques (6 + unfiltered), a total of 70 *stimuli* were presented to each participant.

We expected strong ordering effects, as a participant is likely to become less sensitive to surgery pictures after repeated exposure. Furthermore, a participant is more likely to recognize the content of a picture if it had been presented before unfiltered or with a weaker filter. We addressed this in four ways:

- The order of stimulus presentation was fully randomized across participants, with the constraint that two presentations of the same picture had to be separated by at least two other pictures.
- Experiment instructions warned participants that they would see the same picture multiple times, but asked them to try to answer questions as if they saw each picture for the first time.
- Each stimulus was presented for two seconds only, after which a mask was displayed and the participant was invited to answer the two questions (see Figure 4). Limiting exposure to each picture was expected to slow down habituation. At the same time, pilot studies confirmed that a two-second exposure was more than enough to be able to fully scan and recognize a picture.
- Surgery pictures were interleaved with neutral pictures, which we expected would further slow down habituation, and incite participants to stay focused. Since the alternance of surgery and neutral pictures was purely random, participants could not predict what would come next.

Responses to neutral images were also used to screen participants with poor data. With Likert items, a common and damaging mistake is inversion (e.g., replying "easy" instead of "hard"). Although instructions explicitly warned against this mistake, we *i)* considered a response of 5 or more to the first question for a neutral image as an obvious inversion, and *ii)* discarded the data from all participants who made two or more obvious inversions. This exclusion rule was decided before collecting the data.

The experiment unfolded as follows: first, participants were given an information sheet and a consent form to sign.[5] The information sheet warned participants about the surgery images, with an example, and asked them not to participate if they thought they were hypersensitive. It also informed them that they would be free to stop at any time, should they feel too uncomfortable. Then, participants read instructions on a computer (MacBook Pro 2015 with a 2880 × 1800 retina display and a mouse) and completed the 70 trials. Finally, they were given a research debriefing sheet, a brief questionnaire, and were invited to comment on the experiment. The entire experiment lasted between 15 and 20 minutes.

5.5 Participants

We recruited 30 unpaid participants (9 females, age 21–49, mean = 29, med = 26, SD = 8.6), in conformity with our planned sample size. One additional participant made two obvious inversions as defined in Section 5.4 and was therefore discarded from the analysis. Participants were recruited by email posting to our institution and to students we give classes to, as well as word of mouth to neighbor institutions. Four participants were left-handed and one was ambidextrous. All had normal or corrected-to-normal vision, none were colorblind and they all had higher-education diplomas. Twenty-four reported seeing surgery images or videos before (TV, internet, books), including real surgeries or traumatic injuries (× 3). Eighteen participants participated in a perception study before. Reported sensitivity to surgery images was on average 4.1 (SD = 2.1) on a scale from 1 (not sensitive at all) to 9 (extremely sensitive).

[5]All documents and the experiment software are available at osf.io/4pfes.

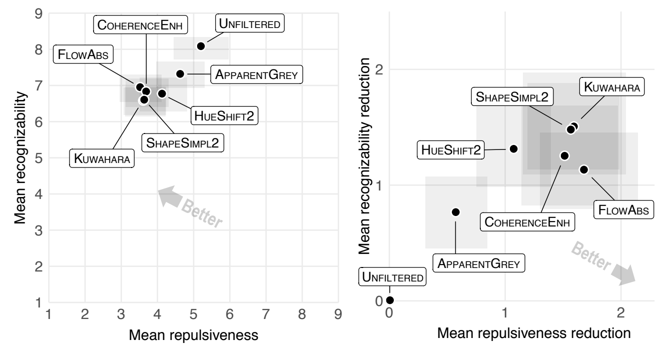


Figure 5: Mean repulsiveness and recognizability ratings for each technique (left); Mean within-subject reduction in repulsiveness and recognizability (right). Boxes are 95% CIs.

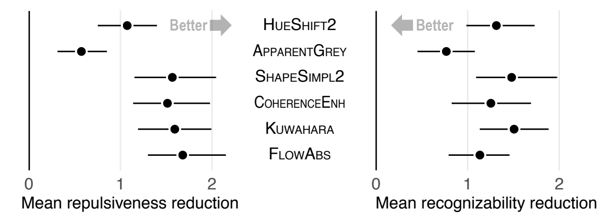


Figure 6: Mean within-subject reduction in repulsiveness and recognizability. Error bars are 95% CIs.

5.6 Results

We report and interpret all our results using interval estimation instead of *p*-values [Cumming 2014; Dragicevic 2016]. All analyses were planned before collecting the data and were preregistered [Cockburn et al. 2018] with the Open Science Framework.[6]

We report on two dependent variables: *repulsiveness*, which is the response to the first question in Figure 4, and *recognizability*, which is the complement ($y = 10 - x$) of the reponse to the second question. For each participant and technique, we averaged responses across all 5 surgery pictures (neutral pictures were not analyzed). Then, for each technique, we derived a point estimate using the mean response across participants, and an interval estimate using the 95% BCa bootstrap confidence interval [Kirby and Gerlanc 2013].

Figure 5-left shows point estimates (dots) and interval estimates (gray boxes) within the full space of possible responses. Roughly speaking, each interval indicates the range of plausible values for the population mean, with the point estimate being about seven times more plausible than the interval endpoints [Cumming 2013]. The mean response for unfiltered surgery images is about 5 on the repulsiveness scale and 8 on the recognizability scale. There is strong evidence that all 6 techniques *i)* yield smaller average recognizability, and *ii)* yield smaller repulsiveness except for ApparentGrey and HueShift2. Unsurprisingly, repulsiveness tends to correlate with recognizability. An ideal processing technique would be an outlier located on the top left of the regression line, but the figure provides no conclusive evidence for such an outlier.

This being a within-subject design, it is more informative to examine average within-subject reduction in repulsiveness and

[6]Preregistered analysis available at osf.io/34vzj.

in recognizability [Cumming 2013], summarized in Figure 5-right. The same results are shown in Figure 6, with a separate plot for each dependent variable. There is overwhelming evidence that all techniques have a within-subject effect on repulsiveness and on recognizability, on average. In terms of repulsiveness reduction, the overlap between error bars [Krzywinski and Altman 2013] suggests that ApparentGrey has the weakest effect, followed by HueShift2, followed by the remaining four. In terms of recognizability reduction the results are less clear, but it is likely that ApparentGrey yields more recognizability than the other techniques.

Overall, all six techniques are effective, but color manipulation appears to be outperformed by space-domain filtering (ShapeSimpl2, CoherenceEnh, Kuwahara, FlowAbs). HueShift2, in particular, is not as effective at making the surgery pictures easier to look at, despite being comparable in terms of preserving informational content. ApparentGrey, on the other hand, simply appears to be a weaker filter: while it does not reduce repulsiveness dramatically, it better preserves image legibility. Among all techniques, FlowAbs and CoherenceEnh offer the best trade-offs if we only consider point estimates, but given the large overlaps in interval estimates the evidence is very weak.

5.7 Participant Feedback

Among the ones who offered feedback, two found FlowAbs to be the best technique. Several participants mentioned that it looked like a cartoon or comic-strip (× 8) making surgery photos easier to look at (× 8). A single participant found that it could make it harder because of the more salient contours, while three participants said that it makes content easier to recognize. HueShift2 was said to make photos easier to look at (× 3) but the content harder to recognize (× 3). Four participants found that it made content appear "unnatural" or "disturbing", while one mentioned that it makes patients look like aliens. The ApparentGrey was deemed easier to look at (× 3) and easier to recognize (× 1), while two participants reported it was harder to recognize the content without colors. Additionally, one participant thought that the ApparentGrey's effectiveness depends on the input picture. This was also reflected by another participant who mentioned that the best technique could be picture-dependent. CoherenceEnh, Kuwahara, and ShapeSimpl2 were reported to be the best technique by one participant each, and to make content harder to recognize by three participants.

6 DISCUSSION AND CONCLUSIONS

In this section, we discuss our results, the limitations of our study, and conclude with future work.

6.1 Summary of findings

We found that all six tested techniques make it easier to look at surgery photos as judged by lay people, even if the effects are moderate and come at a slight cost in legibility. Combining quantitative and qualitative results from both studies, two particularly successful techniques are **FlowAbs** (structure-adaptive filtering [Kyprianidis and Döllner 2008]) and **CoherenceEnh** (coherence-enhancing filtering [Kyprianidis and Kang 2011]). FlowAbs is the only technique integrating edge enhancement, which seems to provide a double benefit: it may increase legibility and recognizability by emphasizing contours, while at the same time producing a cartoon appearance that may make photos look less "real" and therefore less disturbing. Concerning CoherenceEnh and the almost equally successful Kuwahara, ShapeSimpl, and Bilateral, they all provide some form of nonlinear/coherence-enhancing diffusion [Weickert 1999] or anisotropic filtering in order to obtain a kind of a painterly look, although not as explicit as Oilpaint. This subtle painterly look may also make the photos look less real, without compromising too much in legibility. CoherenceEnh was found to be the most legible by surgeons, perhaps because compared to ShapeSimpl, for example, it preserves "*the shape by using a curvature preserving smoothing method that enhances coherence*" [Kyprianidis and Kang 2011] and is more resistant to noise.

The finding that color manipulation was less effective came to us as a surprise, given the emotional force blood appears to carry, and the tradition of altering its color in video games and Japanese animation. Not only HueShift was found to discard information despite the absence of spatial filtering, it was also found to be weaker at affective dampening compared to the approaches mentioned before. ApparentGrey may still have useful applications as a weak surgery filter, as it was found to be relatively usable by surgeons.

Almost all artistic stylization techniques (Watercolor, BrushStrokes, Hatching and Stippling) were found by surgeons to discard too much information to be useful. However, Oilpaint was relatively well received, even if it did not make it in our final experiment. This, together with the fact that FlowAbs is the filter that comes closest to stylization, suggest that stylization still holds promise and deserves to be investigated further.

6.2 Limitations

There are several clear limitations to our experiment. First, our measures are self-reported and may not truly reflect participants' experiences. In particular, we cannot rule out the possibility that responses were tainted by a social desirability bias [Fisher 1993] or, worse, by a good-subject effect [Nichols and Maner 2008]. This could be improved in the future by adopting between-subject designs and/or objective measurements (Section 2.3), but it is unclear whether such methods would provide sufficient sensitivity for populations that are not hypersensitive or BII-phobic. Second, all our results only apply to the surgery pictures we tested. To draw general inferences about the effectiveness of the techniques, one would need to design a study where experimental pictures are randomly drawn from a large collection of representative surgery photos [Judd et al. 2012]. A third important limitation is that we only compared techniques with specific settings, so our findings do not necessarily apply to the techniques in general. Solutions to this include testing many settings, or devising a systematic calibration method that can determine parameters ensuring fair comparisons.

Finally, we only examined 13 techniques and many others remain to be tested. More sophisticated (and possibly more effective) variants exist, especially in the area of color manipulation. Combinations of techniques also merit consideration, such as using edge enhancement from FlowAbs with other techniques, or combining color manipulation with spatial filters. As future work, we also plan to investigate how neural style transfer [Gatys et al. 2016; Johnson et al. 2016; Semmo et al. 2017] can be used to replicate illustrations found in medical textbooks. Such techniques can be quite promising in our context, although their outcome is still hard to control.

6.3 Other Future Work

This article is an initial investigation but opens up exciting avenues for future research. These include supporting surgery videos, other types of medical images than open surgery (e. g., skin diseases), as well as disturbing imagery outside the medical domain, such as offensive user-generated content that can psychologically impact professionals who monitor it [Feinstein et al. 2014]. This topic also raises interesting human-computer interaction questions and opportunities for design, such as the design of adjustable picture censorship in web pages (e. g., Wikipedia) or browsers, or the emotional dampening of live scenes through a "diminished reality" paradigm [Azuma et al. 2001], which if successful could allow general audiences to witness real surgery procedures.

A pending challenge is the use of a standardized picture set for future studies. Even though developments are underway to specify benchmark pictures for expressive rendering [Mould and Rosin 2016, 2017; Rosin et al. 2017], we had to resort to the NAPS and IAPS catalogs, yet even in these sources only few pictures met our requirements. While the selected pictures are available for future work through the named sources, it may be useful to specify a set of standard pictures to study the impact of stylization on affect.

ACKNOWLEDGMENTS

Thanks to Steve Haroz, Yvonne Jansen, Wesley Willett and Pascal Guitton for their help and feedback. This research was partly funded by the Federal Ministry of Education and Research (BMBF), Germany, for the AVA project 01IS15041B.

REFERENCES

Thomas Armstrong, Adam Hemminger, and Bunmi O. Olatunji. 2013. Attentional bias in injection phobia: Overt components, time course, and relation to behavior. *Behaviour Research and Therapy* 51, 6 (June 2013), 266–273. doi: 10.1016/j.brat.2013.02.008

Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. 2001. Recent advances in augmented reality. *IEEE Computer Graphics and Applications* 21, 6 (Nov./Dec. 2001), 34–47. doi: 10.1109/38.963459

Romain Bachy, Jérôme Dias, David Alleysson, and Valérie Bonnardel. 2012. Hue discrimination, unique hues and naming. *Journal of the Optical Society of America A* 29, 2 (Feb. 2012), A60–A68. doi: 10.1364/JOSAA.29.000A60

Charles Bellemare, Luc Bissonnette, and Sabine Kröger. 2014. *Statistical power of within and between-subjects designs in economic experiments.* IZA Discussion Papers, No. 8583. IZA Institute of Labor Economics.

Francesca Benuzzi, Fausta Lui, Davide Duzzi, Paolo F Nichelli, and Carlo A. Porro. 2008. Does it look painful or disgusting? Ask your parietal and cingulate cortex. *Journal of Neuroscience* 28, 4 (Jan. 2008), 923–931. doi: 10.1523/JNEUROSCI.4012-07.2008

Adrien Bousseau, Matt Kaplan, Joëlle Thollot, and François X. Sillion. 2006. Interactive watercolor rendering with temporal coherence and abstraction. In *Proc. NPAR.* ACM, New York, 141–149. doi: 10.1145/1124728.1124751

Martin Čadík. 2008. Perceptual evaluation of color-to-grayscale image conversions. *Computer Graphics Forum* 27, 7 (Oct. 2008), 1745–1754. doi: 10.1111/j.1467-8659.2008.01319.x

John Canny. 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8, 6 (Nov. 1986), 679–698. doi: 10.1109/TPAMI.1986.4767851

Hanah A. Chapman and Adam K. Anderson. 2012. Understanding disgust. *Annals of the New York Academy of Sciences* 1251, 1 (March 2012), 62–76. doi: 10.1111/j.1749-6632.2011.06369.x

Josh M. Cisler, Bunmi O. Olatunji, and Jeffrey M. Lohr. 2009. Disgust, fear, and the anxiety disorders: A critical review. *Clinical Psychology Review* 29, 1 (Feb. 2009), 34–46. doi: 10.1016/j.cpr.2008.09.007

Andy Cockburn, Karl Gutwin, and Alan Dix. 2018. HARK no more: On the preregistration of CHI experiments. In *Proc. CHI.* ACM, New York, Article 141, 12 pages. doi: 10.1145/3173574.3173715

Dorin Comaniciu, Peter Meer, and Senior Member. 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 5 (May 2002), 603–619. doi: 10.1109/34.1000236

Mario Costa Sousa, David S. Ebert, Don Stredney, and Nikolai A. Svakhine. 2005. Illustrative visualization for medical training. In *Proc. CAe.* Eurographics Association, Goslar, Germany, 201–208. doi: 10.2312/COMPAESTH/COMPAESTH05/201-208

Antonio Criminisi, Toby Sharp, Carsten Rother, and Patrick Pérez. 2010. Geodesic image and video editing. *ACM Transactions on Graphics* 29, 5, Article 134 (Oct. 2010), 15 pages. doi: 10.1145/1857907.1857910

Geoff Cumming. 2013. *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* Routledge, New York.

Geoff Cumming. 2014. The new statistics: Why and how. *Psychological Science* 25, 1 (Jan. 2014), 7–29. doi: 10.1177/0956797613504966

Doug DeCarlo and Anthony Santella. 2002. Stylization and abstraction of photographs. *ACM Transactions on Graphics* 21, 3 (July 2002), 769–776. doi: 10.1145/566654.566650

Pierre Dragicevic. 2016. Fair statistical communication in HCI. In *Modern Statistical Methods for HCI*, Judy Robertson and Maurits Kaptein (Eds.). Springer International Publishing, Cham, Switzerland, Chapter 13, 291–330. doi: 10.1007/978-3-319-26633-6_13

Pierre Dragicevic, Wesley Willett, and Tobias Isenberg. 2013. Illustrative data graphic style elements. In *Posters at Expressive.* https://hal.inria.fr/hal-00849081

David J. Duke, Philip J. Barnard, Nick Halper, and Mara Mellin. 2003. Rendering and affect. *Computer Graphics Forum* 22, 3 (Sept. 2003), 359–368. doi: 10.1111/1467-8659.00683

David S. Ebert and Mario Costa Sousa (Eds.). 2006. *Illustrative Visualization for Medicine and Science.* ACM SIGGRAPH 2006 Course Notes, Vol. 6. ACM, New York. doi: 10.1145/1185657.1185691

Falk Eippert, Ralf Veit, Nikolaus Weiskopf, Michael Erb, Niels Birbaumer, and Silke Anders. 2007. Regulation of emotional responses elicited by threat-related stimuli. *Human Brain Mapping* 28, 5 (May 2007), 409–423. doi: 10.1002/hbm.20291

Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. 2008. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics* 27, 3, Article 67 (2008), 10 pages. doi: 10.1145/1360612.1360666

Hasan Sheikh Faridul, Tania Pouli, Christel Chamaret, Jürgen Stauder, Alain Trémeau, and Erik Reinhard. 2014. A survey of color mapping and its applications. In *Eurographics State of the Art Reports.* Eurographics Association, Goslar, Germany, 43–67. doi: 10.2312/egst.20141035

Anthony Feinstein, Blair Audet, and Elizabeth Waknine. 2014. Witnessing images of extreme violence: a psychological study of journalists in the newsroom. *JRSM open* 5, 8 (2014), 2054270414533323. doi: 10.1177/2054270414533323

Robert J. Fisher. 1993. Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research* 20, 2 (Sept. 1993), 303–315. doi: 10.1086/209351

Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2016. Image style transfer using convolutional neural networks. In *Proc. CVPR.* IEEE Computer Society, Los Alamitos, 2414–2423. doi: 10.1109/CVPR.2016.265

Philippe T. Gilchrist and Blaine Ditto. 2012. The effects of blood-draw and injection stimuli on the vasovagal response. *Psychophysiology* 49, 6 (June 2012), 815–820. doi: 10.1111/j.1469-8986.2012.01359.x

Amy A. Gooch, Jeremy Long, Li Ji, Anthony Estey, and Bruce S. Gooch. 2010. Viewing progress in non-photorealistic rendering through Heinlein's lens. In *Proc. NPAR.* ACM, New York, 165–171. doi: 10.1145/1809939.1809959

Amy A. Gooch, Sven C. Olsen, Jack Tumblin, and Bruce Gooch. 2005. Color2Gray: Salience-preserving color removal. *ACM Transactions on Graphics* 24, 3 (July 2005), 634–639. doi: 10.1145/1073204.1073241

Amy A. Gooch and Peter Willemsen. 2002. Evaluating space perception in NPR immersive environments. In *Proc. NPAR.* ACM, New York, 105–110. doi: 10.1145/508530.508549

Bruce Gooch and Amy A. Gooch. 2001. *Non-Photorealistic Rendering.* A K Peters, Ltd., Natick. https://www.taylorfrancis.com/books/9781439864173

Bruce Gooch, Erik Reinhard, and Amy Gooch. 2004. Human facial illustrations: Creation and psychophysical evaluation. *ACM Transactions on Graphics* 23, 1 (2004), 27–44. doi: 10.1145/966131.966133

Mark Grundland and Neil A. Dodgson. 2007. Decolorize: Fast, contrast enhancing, color to grayscale conversion. *Pattern Recognition* 40, 11 (Nov. 2007), 2891–2896. doi: 10.1016/j.patcog.2006.11.003

Anke Haberkamp and Thomas Schmidt. 2014. Enhanced visuomotor processing of phobic images in blood-injury-injection fear. *Journal of Anxiety Disorders* 28, 3 (April 2014), 291–300. doi: 10.1016/j.janxdis.2014.02.001

Peter Hall and Ann-Sophie Lehmann. 2013. Don't measure—Appreciate! NPR seen through the prism of art history. In *Image and Video based Artistic Stylisation*, Paul Rosin and John Collomosse (Eds.). Computational Imaging and Vision, Vol. 42. Springer, London, Heidelberg, Chapter 16, 333–351. doi: 10.1007/978-1-4471-4519-6_16

Nicolas Halper. 2003. *Supportive Presentation for Computer Games.* Ph.D. Dissertation. University of Magdeburg, Germany. urn:nbn:de:101:1-201010122864.

Nick Halper, Mara Mellin, Christoph S. Herrmann, Volker Linneweber, and Thomas Strothotte. 2003. Psychology and non-photorealistic rendering: The beginning of a beautiful relationship. In *Proc. Mensch & Computer: Interaktion in Bewegung*. Teubner Verlag, Stuttgart/Leipzig/Wiesbaden, 277–286. doi: 10.1007/978-3-322-80058-9_28

Robert Hare, Keith Wood, Sue Britain, and Janice Shadman. 1970. Autonomic responses to affective visual stimulation. *Psychophysiology* 7, 3 (Nov. 1970), 408–417. doi: 10.1111/j.1469-8986.1970.tb01766.x

Kaiming He, Jian Sun, and Xiaoou Tang. 2013. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 6 (June 2013), 1397–1409. doi: 10.1109/TPAMI.2012.213

Aaron Hertzmann. 1998. Painterly rendering with curved brush strokes of multiple sizes. In *Proc. SIGGRAPH*. ACM, New York, 453–460. doi: 10.1145/280814.280951

Tobias Isenberg. 2013. Evaluating and validating non-photorealistic and illustrative rendering. In *Image and Video based Artistic Stylisation*, Paul Rosin and John Collomosse (Eds.). Computational Imaging and Vision, Vol. 42. Springer, London, Heidelberg, Chapter 15, 311–331. doi: 10.1007/978-1-4471-4519-6_15

Tobias Isenberg. 2016. Interactive NPAR: What type of tools should we create? In *Proc. NPAR*. Eurographics Association, Goslar, Germany, 89–96. doi: 10.2312/exp.20161067

Rong Jin and Luo Si. 2004. A study of methods for normalizing user ratings in collaborative filtering. In *Proc. SIGIR*. ACM, New York, 568–569. doi: 10.1145/1008992.1009124

Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In *Proc. ECCV*. Springer International, Cham, Switzerland, 694–711. doi: 10.1007/978-3-319-46475-6_43

Charles M. Judd, Jacob Westfall, and David A. Kenny. 2012. Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology* 103, 1 (July 2012), 54–69. doi: 10.1037/a0028347

Henry Kang and Seungyong Lee. 2008. Shape-simplifying image abstraction. *Computer Graphics Forum* 27, 7 (Oct. 2008), 1773–1780. doi: 10.1111/j.1467-8659.2008.01322.x

Henry Kang, Seungyong Lee, and Charles K. Chui. 2007. Coherent line drawing. In *Proc. NPAR*. ACM, New York, 43–50. doi: 10.1145/1274871.1274878

Henry Kang, Seungyong Lee, and Charles K. Chui. 2009. Flow-based image abstraction. *IEEE Transactions on Visualization and Computer Graphics* 15, 1 (Jan./Feb. 2009), 62–76. doi: 10.1109/TVCG.2008.81

Michael Kass and Justin Solomon. 2010. Smoothed local histogram filters. *ACM Transactions on Graphics* 29, 4, Article 100 (July 2010), 10 pages. doi: 10.1145/1833351.1778837

Kill Bill Wiki Contributors. 2017. Crazy 88. (October 2017). http://killbill.wikia.com/wiki/Crazy_88

Yongjin Kim, Cheolhun Jang, Julien Demouth, and Seungyong Lee. 2009. Robust color-to-gray via nonlinear global mapping. *ACM Transactions on Graphics* 28, 5, Article 161 (Dec. 2009), 4 pages. doi: 10.1145/1618452.1618507

Kris N. Kirby and Daniel Gerlanc. 2013. BootES: An R package for bootstrap confidence intervals on effect sizes. *Behavior Research Methods* 45, 4 (Dec. 2013), 905–927. doi: 10.3758/s13428-013-0330-5

Rafael Klorman, Roger P. Weissberg, and Alan R. Wiesenfeld. 1977. Individual differences in fear and autonomic reactions to affective stimulation. *Psychophysiology* 14, 1 (Jan. 1977), 45–51. doi: 10.1111/j.1469-8986.1977.tb01154.x

Martin Krzywinski and Naomi Altman. 2013. Points of significance: Error bars. *Nature Methods* 10, 10 (Oct. 2013), 921–922. doi: 10.1038/nmeth.2659

M. Kuwahara, K. Hachimura, S. Eiho, and M. Kinoshita. 1976. Processing of RI-angiocardiographic images. In *Digital Processing of Biomedical Images*. Plenum Press, New York, 187–202. doi: 10.1007/978-1-4684-0769-3_13

Jan Eric Kyprianidis. 2011. Image and video abstraction by multi-scale anisotropic Kuwahara filtering. In *Proc. NPAR*. ACM, New York, 55–64. doi: 10.1145/2024676.2024686

Jan Eric Kyprianidis, John Collomosse, Tinghuai Wang, and Tobias Isenberg. 2013. State of the “art”: A taxonomy of artistic stylization techniques for images and video. *IEEE Transactions on Visualization and Computer Graphics* 19, 5 (May 2013), 866–885. doi: 10.1109/TVCG.2012.160

Jan Eric Kyprianidis and Jürgen Döllner. 2008. Image abstraction by structure adaptive filtering. In *Proc. EG UK—Theory and Practice of Computer Graphics*. The Eurographics Association, Goslar, Germany, 51–58. doi: 10.2312/LocalChapterEvents/TPCG/TPCG08/051-058

Jan Eric Kyprianidis and Henry Kang. 2011. Image and video abstraction by coherence-enhancing filtering. *Computer Graphics Forum* 30, 2 (April 2011), 593–602. doi: 10.1111/j.1467-8659.2011.01882.x

Jan Eric Kyprianidis, Henry Kang, and Jürgen Döllner. 2009. Image and video abstraction by anisotropic Kuwahara filtering. *Computer Graphics Forum* 28, 7 (Oct. 2009), 1955–1963. doi: 10.1111/j.1467-8659.2009.01574.x

Peter J. Lang, Margaret M. Bradley, and Bruce N. Cuthbert. 1997. International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention* (1997), 39–58.

Peter J. Lang, Mark K. Greenwald, Margaret M. Bradley, and Alfons O. Hamm. 1993. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology* 30, 3 (May 1993), 261–273. doi: 10.1111/j.1469-8986.1993.tb03352.x

Mark A. Lumley and Barbara G. Melamed. 1992. Blood phobics and nonphobics: Psychological differences and affect during exposure. *Behaviour Research and Therapy* 30, 5 (Sept. 1992), 425–434. doi: 10.1016/0005-7967(92)90026-D

Kede Ma, Tiesong Zhao, Kai Zeng, and Zhou Wang. 2015. Objective quality assessment for color-to-gray image conversion. *IEEE Transactions on Image Processing* 24, 12 (Dec. 2015), 4673–4685. doi: 10.1109/TIP.2015.2460015

Regan L. Mandryk, David Mould, and Hua Li. 2011. Evaluation of emotional response to non-photorealistic images. In *Proc. NPAR*. ACM, New York, 7–16. doi: 10.1145/2024676.2024678

Artur Marchewka, Łukasz Żurawski, Katarzyna Jednoróg, and Anna Grabowska. 2014. The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods* 46, 2 (June 2014), 596–610. doi: 10.3758/s13428-013-0379-1

D. Marr and E Hildreth. 1980. Theory of edge detection. *Proceedings of the Royal Society B* 207, 1167 (Feb. 1980), 187–217. doi: 10.1098/rspb.1980.0020

Domingo Martín, Germán Arroyo, M. Victoria Luzón, and Tobias Isenberg. 2010. Example-based stippling using a scale-dependent grayscale process. In *Proc. NPAR*. ACM, New York, 51–61. doi: 10.1145/1809939.1809946

Domingo Martín, Germán Arroyo, M. Victoria Luzón, and Tobias Isenberg. 2011. Scale-dependent and example-based stippling. *Computers & Graphics* 35, 1 (Feb. 2011), 160–174. doi: 10.1016/j.cag.2010.11.006

Rachel McDonnell, Martin Breidt, and Heinrich H. Bülthoff. 2012. Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics* 31, 4, Article 91 (July 2012), 11 pages. doi: 10.1145/2185520.2185587

David Mould. 2012. Texture-preserving abstraction. In *Proc. NPAR*. Eurographics Association, Goslar, Germany, 75–82. doi: 10.2312/PE/NPAR/NPAR12/075-082

David Mould. 2014. Authorial subjective evaluation of non-photorealistic images. In *Proc. NPAR*. ACM, New York, 49–56. doi: 10.1145/2630397.2630400

David Mould, Regan L. Mandryk, and Hua Li. 2012. Emotional response and visual attention to non-photorealistic images. *Computers & Graphics* 36, 6 (Oct. 2012), 658–672. doi: 10.1016/j.cag.2012.03.039

David Mould and Paul L. Rosin. 2016. A benchmark image set for evaluating stylization. In *Proc. NPAR*. The Eurographics Association, Goslar, Germany, 11–20. doi: 10.2312/exp.20161059

David Mould and Paul L. Rosin. 2017. Developing and applying a benchmark for evaluating image stylization. *Computers & Graphics* 67 (2017), 58–76. doi: 10.1016/j.cag.2017.05.025

László Neumann, Martin Čadík, and Antal Nemcsics. 2007. An efficient perception-based adaptive color to gray transformation. In *Proc. CAe*. Eurographics Association, Goslar, Germany, 73–80. doi: 10.2312/COMPAESTH/COMPAESTH07/073-080

Austin Lee Nichols and Jon K. Maner. 2008. The good-subject effect: Investigating participant demand characteristics. *The Journal of General Psychology* 135, 2 (2008), 151–166. doi: 10.3200/GENP.135.2.151-166

Bunmi O. Olatunji, Josh Cisler, Dean McKay, and Mary L. Phillips. 2010. Is disgust associated with psychopathology? Emerging research in the anxiety disorders. *Psychiatry Research* 175, 1–2 (Jan. 2010), 1–10. doi: 10.1016/j.psychres.2009.04.007

Bunmi O. Olatunji, Jonathan Haidt, Dean McKay, and Bieke David. 2008. Core, animal reminder, and contamination disgust: Three kinds of disgust with distinct personality, behavioral, physiological, and clinical correlates. *Journal of Research in Personality* 42, 5 (Oct. 2008), 1243–1259. doi: 10.1016/j.jrp.2008.03.009

Lars-Göran Öst, Ulf Sterner, and Inga-Lena Lindahl. 1984. Physiological responses in blood phobics. *Behaviour Research and Therapy* 22, 2 (1984), 109–117. doi: 10.1016/0005-7967(84)90099-8

Giuseppe Papari, Nicolai Petkov, and Patrizio Campisi. 2007. Artistic edge and corner enhancing smoothing. *IEEE Transactions on Image Processing* 16, 10 (Oct. 2007), 2449–2662. doi: 10.1109/TIP.2007.903912

Charles Perin, Pierre Dragicevic, and Jean-Daniel Fekete. 2014. Revisiting Bertin matrices: New interactions for crafting tabular visualizations. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (Dec. 2014), 2082–2091. doi: 10.1109/TVCG.2014.2346279

William K. Pratt. 2007. *Digital Image Processing: PIKS Scientific Inside* (4th ed.). John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/0470097434

Emil Praun, Hughes Hoppe, Matthew Webb, and Adam Finkelstein. 2001. Real-time hatching. In *Proc. SIGGRAPH*. ACM, New York, 581–586. doi: 10.1145/383259.383328

Bernhard Preim and Charl Botha. 2014. Illustrative medical visualization. In *Visual Computing for Medicine* (2nd ed.). Morgan Kaufmann, Boston, Chapter 12, 451–508. doi: 10.1016/B978-0-12-415873-3.00012-2

Peter Rautek, Stefan Bruckner, Eduard Gröller, and Ivan Viola. 2008. Illustrative visualization: New technology or useless tautology? *ACM SIGGRAPH Computer Graphics* 42, 3, Article 4 (Aug. 2008), 8 pages. doi: 10.1145/1408626.1408633

Sonja Rohrmann and Henrik Hopp. 2008. Cardiovascular indicators of disgust. *International Journal of Psychophysiology* 68, 3 (June 2008), 201–208. doi: 10.1016/j.ijpsycho.2008.01.011

Paul Rosin and John Collomosse (Eds.). 2013. *Image and Video based Artistic Stylisation*. Computational Imaging and Vision, Vol. 42. Springer, London, Heidelberg. doi: 10.1007/978-1-4471-4519-6

Paul L. Rosin, David Mould, Itamar Berger, John Collomosse, Yu-Kun Lai, Chuan Li, Hua Li, Ariel Shamir, Michael Wand, Tinghuai Wang, and Holger Winnemöller. 2017. Benchmarking non-photorealistic rendering of portraits. In *Proc. NPAR*. ACM, New York, Article 11, 12 pages. doi: 10.1145/3092919.3092921

David H. Salesin. 2002. Non-Photorealistic Animation & Rendering: 7 Grand Challenges. Keynote talk at NPAR. (June 2002).

Craig N. Sawchuk, Jeffrey M. Lohr, David H. Westendorf, Suzanne A. Meunier, and David F. Tolin. 2002. Emotional responding to fearful and disgusting stimuli in specific phobics. *Behaviour Research and Therapy* 40, 9 (Sept. 2002), 1031–1046. doi: 10.1016/S0005-7967(01)00093-6

Anne Schienle, R. Stark, B. Walter, C. Blecker, U. Ott, P. Kirsch, G. Sammer, and D. Vaitl. 2002. The insula is not specifically involved in disgust processing: An fMRI study. *NeuroReport* 13, 16 (Nov. 2002), 2023–2026. doi: 10.1097/00001756-200211150-00006

Jutta Schumann, Thomas Strothotte, Andreas Raab, and Stefan Laser. 1996. Assessing the effect of non-photorealistic rendered images in CAD. In *Proc. CHI*. ACM, New York, 35–42. doi: 10.1145/238386.238398

Amir Semmo, Tobias Isenberg, and Jürgen Döllner. 2017. Neural style transfer: A paradigm shift for image-based artistic rendering?. In *Proc. NPAR*. ACM, New York, Article 5, 13 pages. doi: 10.1145/3092919.3092920

Amir Semmo, Daniel Limberger, Jan Eric Kyprianidis, and Jürgen Döllner. 2016. Image stylization by interactive oil paint filtering. *Computers & Graphics* 55 (April 2016), 157–171. doi: 10.1016/j.cag.2015.12.001

Maria Shugrina, Margrit Betke, and John Collomosse. 2006. Empathic painting: Interactive stylization through observed emotional state. In *Proc. NPAR*. ACM, New York, 87–96. doi: 10.1145/1124728.1124744

Kaleigh Smith, Pierre-Edouard Landes, Joëlle Thollot, and Karol Myszkowski. 2008. Apparent greyscale: A simple and fast conversion to perceptually accurate images and video. *Computer Graphics Forum* 27, 2 (April 2008), 193–200. doi: 10.1111/j.1467-8659.2008.01116.x

Thomas Strothotte and Stefan Schlechtweg. 2002. *Non-Photorealistic Computer Graphics. Modeling, Animation, and Rendering*. Morgan Kaufmann Publishers, San Francisco. doi: 10.1016/B978-1-55860-787-3.50019-0

David F. Tolin, Jeffrey M. Lohr, Craig N. Sawchuk, and Thomas C. Lee. 1997. Disgust and disgust sensitivity in blood-injection-injury and spider phobia. *Behaviour Research and Therapy* 35, 10 (Oct. 1997), 949–953. doi: 10.1016/S0005-7967(97)00048-X

Carlo Tomasi and Roberto Manduchi. 1998. Bilateral filtering for gray and color images. In *Proc. ICCV*. IEEE Computer Society, Los Alamitos, 839–846. doi: 10.1109/ICCV.1998.710815

Roger Tooth. 2014. Graphic content: When photographs of carnage are too upsetting to publish. *The Guardian* (July 2014). https://www.theguardian.com/world/2014/jul/23/graphic-content-photographs-too-upsetting-to-publish-gaza-mh17-ukraine

TVTropes Contributors. 2018a. AdjustableCensorship. (March 2018). http://tvtropes.org/pmwiki/pmwiki.php/Main/AdjustableCensorship

TVTropes Contributors. 2018b. BlackBlood. (March 2018). http://tvtropes.org/pmwiki/pmwiki.php/Main/BlackBlood

Ivan Viola, Mario Costa Sousa, David S. Ebert, Bill Andrews, Bruce Gooch, and Christian Tietjen. 2006. Illustrative visualization for medicine and science. In *Eurographics Tutorials*. Eurographics Association, Goslar, Germany, 1061–1200. doi: 10.2312/egt.20061068

Ivan Viola and Tobias Isenberg. 2018. Pondering the concept of abstraction in (illustrative) visualization. *IEEE Transactions on Visualization and Computer Graphics* 24 (2018). doi: 10.1109/TVCG.2017.2747545 To appear.

Miaoyi Wang, Bin Wang, Yun Fei, Kanglai Qian, Wenping Wang, Jiating Chen, and Jun-Hai Yong. 2014. Towards photo watercolorization with artistic verisimilitude. *IEEE Transactions on Visualization and Computer Graphics* 20, 10 (Feb. 2014), 1451–1460. doi: 10.1109/TVCG.2014.2303984

Joachim Weickert. 1999. Coherence-enhancing diffusion filtering. *International Journal of Computer Vision* 31, 2–3 (April 1999), 111–127. doi: 10.1023/A:1008009714131

Wikipedia Contributors. 2010a. QA Wikimedia Commons images review. (2010). https://wikimediafoundation.org/wiki/QA_Wikimedia_Commons_images_review,_May_2010

Wikipedia Contributors. 2010b. Wikipedia:Sexual content/FAQ. (2010). https://en.wikipedia.org/wiki/Wikipedia:Sexual_content/FAQ

Holger Winnemöller, Jan Eric Kyprianidis, and Sven C. Olsen. 2012. XDoG: An extended difference-of-Gaussians compendium including advanced image stylization. *Computers & Graphics* 36, 6 (2012), 740–753. doi: 10.1016/j.cag.2012.03.004

Holger Winnemöller, Sven C. Olsen, and Bruce Gooch. 2006. Real-time video abstraction. *ACM Transactions on Graphics* 25, 3 (July 2006), 1221–1226. doi: 10.1145/1141911.1142018

Jana Wrase, Sabine Klein, Sabine M. Gruesser, Derik Hermann, Herta Flor, Karl Mann, Dieter F. Braus, and Andreas Heinz. 2003. Gender differences in the processing of standardized emotional visual stimuli in humans: A functional magnetic resonance imaging study. *Neuroscience Letters* 348, 1 (Sept. 2003), 41–45. doi: 10.1016/S0304-3940(03)00565-2

Li Xu, Cewu Lu, Yi Xu, and Jiaya Jia. 2011. Image smoothing via L_0 gradient minimization. *ACM Transactions on Graphics* 30, 6, Article 174 (Dec. 2011), 12 pages. doi: 10.1145/2070781.2024208

ZeldaWiki Contributors. 2018. Controversy in The Legend of Zelda Series. (March 2018). https://zelda.gamepedia.com/Controversy_in_The_Legend_of_Zelda_Series